第27卷 第6期 工 程 数 学 学 报 Vol. 27 No. 6
2010年12月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Dec. 2010

文章编号:1005-3085(2010)06-0986-09

噪声为厚尾过程的非参数函数变点的小波估计*

李晓燕[1]， 田 铮[1,2]， 齐培艳[1]， 赵文芝[1]

(1- 西北工业大学应用数学系，西安 710072; 2- 西北工业大学计算机科学技术系，西安 710072)

摘 要: 本文给出了随机设计下非参数回归模型中噪声为无穷方差过程的小波检测和估计方法。利用基于经验小波系数的检验统计量，在原假设成立的条件下，推导出任意尺度上检验的临界值，证明了检验的一致性；在备择假设成立的条件下，得到变点个数、变点位置的相合估计与收敛速度。数值模拟以及IBM股票数据实例分析的结果均表明方法是有效的。

关键词: 随机设计；非参数回归模型；无穷方差过程；变点；小波估计

分类号: **AMS**(2000) 60E05; 62E20; 62H15 **中图分类号**: O212.1 **文献标识码**: A

1 引言

变点分析不仅是统计学和时间序列分析研究的热点理论问题之一，而且其理论研究成果在诸多领域都有应用，如：经济、金融、医学和气象学等领域。近年来关于变点成果的综述可见文献[1,2]。

非参数回归模型能够很好地描述工程系统中的非线性关系，具有广泛的研究背景[3-6]。小波变换能够自动改变窗长，因此可以很好地反映函数的整体和局部特征。这一特性使得小波成为处理非参数变点问题的有力工具。Wang[3] 基于小波系数绝对值的检验统计量对变点进行了检测和估计；Odgen[4] 利用小波系数的平方累积和对均值变点进行了检测；Raimondo[5,6] 由小波给出了非参数回归函数的minimax估计，并解决了方差有穷厚尾信号的变点检验问题。

文献[3-6]考虑的都是噪声为方差有穷序列的固定设计情形，而近年来引起金融与经济研究领域中学者的重视的是：厚尾随机序列能够描述许多金融资产收益率分布中正态分布无法描述的尖峰、厚尾等特性[7]等。基于此，本文利用小波方法研究随机设计下噪声为无穷方差厚尾过程的非参数模型中，回归函数变点的小波检测和估计问题。第2节给出本文的模型与假设，第3节与第4节是主要研究成果，最后是数值模拟与实例分析。

2 模型与假设

设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 是从如下随机设计回归模型获得的观察值序列

$$Y_i=f(X_i)+\varepsilon_i,\quad i=1,2,\cdots,n. \tag{1}$$

其中 $\{\varepsilon_i,i=1,2,\cdots,n\}$ 为中心化的i.i.d.噪声序列，$X_i\sim U[0,1]$，f 的支撑集为[0,1]。本文假设条件如下：

收稿日期: 2008-01-21. **作者简介**: 李晓燕(1984年12月生)，女，硕士. 研究方向：信息处理与时间序列.

***基金项目**: 国家自然科学基金(60972150; 10926197).

假设 1 随机变量序列 $\{\varepsilon_i,\ i=1,2,\cdots,n\}$ 具有一维对称边缘分布且 ε_1 的特征函数满足

$$E\exp\{\mathrm{i}u\varepsilon_1\}=\exp\{-|u|^{\kappa}\},\quad 1<\kappa<2. \tag{2}$$

假设 2 f 在 [0,1] 上有界。

注 1 ε_1 的特征函数为稳定分布的特征函数，即 ε_1 属于特征指数为 κ 的稳定分布的吸收域[8]。当 $1<\kappa<2$ 时，ε_1 的方差不存在，为无穷方差厚尾随机变量。

本文研究如下假设检验问题：

H_0: f 是光滑函数，在 $[0,1]$ 上至少连续可微；

$H_1(m)$: f 中存在 q 个跳跃点，且 $1\le q\le m$ (已知)，除 q 个跳跃点之外 f 是光滑的。

对样本 $\{X_1,\cdots,X_n\}$ 进行排序得到次序统计量 $X_{(1)}\le\cdots\le X_{(n)}$，此时模型 (1)变为

$$Y_{[i]}=f(X_{(i)})+e_i,\quad i=1,2,\cdots,n. \tag{3}$$

其中 $Y_{[i]}$ 和 e_i 分别是 Y_i 和 ε_i 的重新排列。模型 (3) 又等价于

$$Y_{[i]}=f(i/n)+\big[f(X_{(i)})-f(i/n)\big]+e_i=f(i/n)+\xi_i+e_i, \tag{4}$$

其中 $\xi_i=f(X_{(i)})-f(i/n)$。

3 小波变换与变点检测

令 $\psi(u)$ 是小波母函数，通过伸缩和平移得到一列小波基函数

$$\psi_{j,k}(u)=2^{j/2}\psi(2^ju-k),\quad j\in N,\quad k\in Z.$$

函数 g 经过连续小波变换得到的小波系数为

$$(w_{a,b})=\int g(u)\psi_{a,b}(u)\,\mathrm{d}u,\quad a\in N,\quad b\in Z.$$

实际中通常在离散时间点上观测函数，因此小波变换的离散化是十分必要的。当给定离散数据 $Y=(Y_1,\cdots,Y_n)$ 时，Mallat[9] 提出的塔型算法可快速计算 Y 经过离散小波变换得到的经验小波系数 $(w_{j,k})=WY$，其中 W 是依赖于小波基的一个正交矩阵，指标 j 为分辨水平 (对应着频率 $\omega=2^{-j}$)，满足 $0\le j\le J$, $2^J=n$，指标 k 为时间 (空间) 参数，对应着时间位置 $t=k/2^j$。

考虑到经验小波变换的 L^2-范数，Härdle[10] 用以下和式来近似小波变换的积分式

$$w_{j,k}(g)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(i/n)g(i/n).$$

故对任意的 j, k，模型 (4) 经过小波变换并用上述方法近似有

$$w_{j,k}=w_{j,k}(f)+w_{j,k}(\xi)+w_{j,k}(e), \tag{5}$$

其中

$$w_{j,k}(f)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(i/n)f(i/n),\quad w_{j,k}(\xi)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}\xi_i,\quad w_{j,k}(e)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(i/n)e_i.$$

本文选取 Haar 小波，则 $w_{j,k}(f)$，$w_{j,k}(\xi)$，$w_{j,k}(e)$ 分别有如下性质：

引理1[6] 当原假设 H_0 成立时，f 为连续可微函数，对所有的分辨水平 $j \geq 3$，以及任意的 k，$k = 0, 1, \cdots, 2^j - 1$，有 $|w_{j,k}(f)| \leq c_1\sqrt{n2^{-3j}}$；当备择假设 H_1 成立时，f 中至少存在一个跳跃点 $x \in [k/2^j, (k+1)/2^j)$，使得 $|w_{j,k}(f)| \geq c_2\sqrt{n2^{-j}}$，其中 c_1, c_2 均为只依赖于 f 的常数，为方便表述取 $c_1 = c_2 = 1$。

引理2 在任意水平 j 下，对所有的 $k \in \{0, \cdots, 2^j - 1\}$，有

$$w_{j,k}(\xi) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(i/n)\big(f(X_{(i)}) - f(i/n)\big) = o_P(1).$$

证明 因为 $X_i \sim U[0,1]$，所以 $X_i \sim \mathrm{Beta}(i, n-i+1)$，故 $|X_{(i)} - i/n| = O_P(1/n)$，再由假设2，有

$$w_{j,k}(\xi) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(i/n)\big(f(X_{(i)}) - f(i/n)\big) = O_P\Big(\frac{1}{\sqrt{n}}\Big) = o_P(1).$$

对于 $w_{j,k}(e)$，考虑到其性质的证明过程，令

$$w_k = \frac{\sqrt{n}}{n^{1/\kappa}}w_{j,k} = w_k(f) + w_k(\xi) + w_k(e),$$

其中

$$w_k(f) = \frac{\sqrt{n}}{n^{1/\kappa}}w_{j,k}(f), \quad w_k(\xi) = \frac{\sqrt{n}}{n^{1/\kappa}}w_{j,k}(\xi), \quad w_k(e) = \frac{\sqrt{n}}{n^{1/\kappa}}w_{j,k}(e).$$

引理3 在任意分辨水平 j 下，对所有的 $k \in \{0, \cdots, 2^j - 1\}$ 有

$$w_k(e) \stackrel{d}{\longrightarrow} 2^{j(1/2-1/\kappa)}z, \quad n \longrightarrow \infty,$$

其中随机变量 z 满足假设1。

证明 利用 Haar 小波的性质以及假设1可得

$$w_k(e) = \frac{1}{n^{1/\kappa}}\sum_{i=1}^{n}\psi_{j,k}(i/n)e_i = \frac{1}{n^{1/\kappa}}\sum_{i=1}^{n/2^j}2^{j/2}(\pm)e_i \stackrel{d}{=} \frac{1}{(n/2^j)^{1/\kappa}}2^{j(1/2-1/\kappa)}\sum_{i=1}^{n/2^j}e_i,$$

这里 $\xi \stackrel{d}{=} \eta$ 表示随机变量 ξ 与 η 依分布相等。由广义中心极限定理[8] 知，若 $\varepsilon_1, \cdots, \varepsilon_n$ 是 i.i.d. 随机变量序列，ε_1 的特征函数为 (2) 式所定义，则有

$$\frac{1}{n^{1/\kappa}}(\varepsilon_1 + \cdots + \varepsilon_n) \stackrel{d}{\longrightarrow} z,$$

这里随机变量 z 满足假设1，并且 z 与 ε_1 的特征指数相同，因此

$$w_k(e) \stackrel{d}{\longrightarrow} 2^{j(1/2-1/\kappa)}z, \quad n \longrightarrow \infty.$$

3.1 小波系数阈值与检验临界值的选取

当原假设 H_0 成立时，小波系数的大小主要由 $w_k(e)$ 决定；而当备择假设 H_1 成立时，f 中存在跳跃点，在跳跃点附近小波系数的绝对值很大[3]。因此，本文对分辨水平 j 施加一定的条件以保证至少存在有限个小波系数大于给定的阈值 $v_n = (2^{-3j})^{\frac{1}{2}}n^{1-\frac{1}{\kappa}}$，即有：

定理1 在水平 j 下，对小波系数 $|w_k|$ 进行排序得到 $|w_{(1)}| \geq |w_{(2)}| \geq \cdots \geq |w_{(2^j)}|$，当原假设 H_0 成立时，如果分辨水平满足

$$2^j \asymp n^\delta, \quad \delta > \frac{\kappa-1}{2\kappa}, \quad 1 < \kappa < 2, \tag{6}$$

则对任意的常数$m \geq 1$，有

$$P\big\{|w_{(m)}| > v_n\big\} \longrightarrow 1, \quad n \longrightarrow \infty.$$

这里$a_n \asymp b_n$指的是存在常数c_3, c_4，使得对足够大的n有$c_3 b_n \leq a_n \leq c_4 b_n$。

证明　只考虑$m = 1$的情形，$m > 1$可类似处理。由Haar小波的性质可知$w_k, k = 0, 1, \cdots, 2^j - 1$是独立的随机变量序列，因此

$$P\big\{|w_{(1)}| \leq v_n\big\} = P\Big\{\max_{k=0,1,\cdots,2^j-1} |w_k| \leq v_n\Big\} = \prod_{k=0}^{2^j-1} P\big\{|w_k| \leq v_n\big\}.$$

当$0 \leq x \leq 1$时，由不等式$1 - x \leq \exp(-x)$可知

$$P\big\{|w_{(1)}| \leq v_n\big\} \leq \exp\Big\{-\sum_{k=0}^{2^j-1} P\big(|w_k| > v_n\big)\Big\}.$$

因此只需证明当$n \longrightarrow \infty$时

$$A_n = \sum_{k=0}^{2^j-1} P\big(|w_k| > v_n\big) \longrightarrow \infty$$

成立。

当原假设H_0成立时，有$|w_k(f)| \leq v_n$，因此$|w_k| \geq |w_k(\xi) + w_k(e)| - v_n$，从而由引理2知

$$\begin{aligned} P\big\{|w_k| > v_n\big\} &\geq P\big\{|w_k(\xi) + w_k(e)| > 2v_n\big\} = P\big\{|w_k(e)| > 2v_n\big\} \\ &= P\Big\{\Big|\frac{1}{(n/2^j)^{1/\kappa}} \sum_{i=1}^{n/2^j} e_i\Big| > 2 \cdot 2^{-j(\frac{1}{2}-\frac{1}{\kappa})} v_n\Big\}. \end{aligned}$$

再由引理3、假设1和稳定分布的尾概率性质可得，当$n \longrightarrow \infty$时

$$P\big\{|w_k| > v_n\big\} = 2P\big\{z > 2 \cdot 2^{-j(\frac{1}{2}-\frac{1}{\kappa})} v_n\big\} \sim 2C_\kappa \big\{2 \cdot 2^{-j(\frac{1}{2}-\frac{1}{\kappa})} v_n\big\}^{-\kappa} = C 2^{j(2\kappa-1)} n^{1-\kappa},$$

其中$h(x) \sim g(x)$表示当$x \longrightarrow \infty$时，$\lim\limits_{x \longrightarrow \infty} h(x)/g(x) = 1$，随机变量$z$同引理3所指，

$$C_\kappa = \Big(2\int_x^\infty x^{-\kappa} \sin\kappa \,\mathrm{d}x\Big)^{-1} = \frac{1}{\pi}\Gamma(\kappa) \sin\Big(\frac{\pi\kappa}{2}\Big),$$

C为常数，下文各式中可能取值不同。故在(6)下

$$A_n = \sum_{k=0}^{2^j-1} P\big(|w_k| > v_n\big) \geq C 2^j 2^{j(2\kappa-1)} n^{1-\kappa} \longrightarrow \infty.$$

证毕

定理1表明：噪声小波系数$|w_k|$至少有m个大于阈值v_n，因此，可能由于v_n设置的偏小在实际中不实用，故本文构造检验统计量

$$N_i = |w_{(i)}| - |w_{(m+1)}|, \quad i = 1, 2, \cdots, m. \tag{7}$$

再结合定理1，给出检验的$100(1-\beta)\%$置信界。

定理2 令$0<\beta<1$为任意常数

$$c_i=\Big(\frac{2^{j(1-\kappa/2)}\beta}{2\nu_\kappa m(m-i+1)}\Big)^{-1/\kappa},\quad R_n(\beta)=\cup_{i=1,\cdots,m}(N_i>c_i), \tag{8}$$

其中

$$\nu_\kappa=\frac{1}{\pi}\Gamma(\kappa)\sin\left(\frac{\kappa}{2}\right),\quad 1<\kappa<2,$$

则当原假设H_0成立且水平j满足(6)式时，有$P\{R_n(\beta)\}\le\beta$, $n\longrightarrow\infty$。

证明 令$B_n=\{|w_{(m+1)}|>v_n\}$，则对任意的x，有

$$P\big(N_i>x\,|\,B_n\big)\le P\big(|w_{(i)}|-v_n>x\,|\,B_n\big),$$

记$|w_{(i)}|\ge|w_{(i+1)}|\ge\cdots\ge|w_{(m)}|$在排序之前的小波系数为$|w_i|,|w_{i+1}|,\cdots,|w_m|$，由引理1

$$\begin{aligned}P\big\{|w_{(i)}|-v_n>x\mid B_n\big\} &= P\Big\{\max_{k=i,\cdots,m}|w_k|-v_n>x\mid B_n\Big\}\\ &\le\sum_{k=i}^{m}P\big\{|w_k|-v_n>x\mid B_n\big\}\le\sum_{k=i}^{m}P\big\{|w_k(\xi)+w_k(e)|>x\mid B_n\big\}.\end{aligned}$$

再由定理1、引理2、引理3、假设1和稳定分布的尾概率性质有

$$\begin{aligned}\lim_{n\longrightarrow\infty}P\{N_i>x\mid B_n\} &\le \lim_{n\longrightarrow\infty}\sum_{k=i}^{m}P\big\{|w_k(\xi)+w_k(e)|>x\mid B_n\big\}\\ &= \lim_{n\longrightarrow\infty}\sum_{k=i}^{m}P\big\{|w_k(e)|>x\big\}\sim2(m-i+1)C_k\Big(2^{-j(\frac{1}{2}-\frac{1}{\kappa})}x\Big)^{-\kappa}.\end{aligned}$$

取$x=c_i$，则

$$\lim_{n\longrightarrow\infty}P\{R_n(\beta)\}=\lim_{n\longrightarrow\infty}\sum_{i=1}^{m}P\big\{N_i>x\mid B_n\big\}\le2\sum_{i=1}^{m}(m-i+1)C_k2^{-j(1-\frac{\kappa}{2})}c_i^{-\kappa}=\beta.$$

证毕

注2 Wang[3]对多变点进行检验时所用临界值只有一个，而本文的临界值随变点个数的变换而变化，这是因为噪声小波系数的前m个最大值的分布不同。

注3 如果所讨论的噪声项具有尺度因子$\lambda>0$，即$Y_i=f(X_i)+\lambda\varepsilon_i$, $i=1,2,\cdots,n$，则临界值应为λc_i。

3.2 检验的一致性

定理3 当备择假设$H_1(m)$成立时，如果水平j满足(6)式，则有$\lim\limits_{n\longrightarrow\infty}P\{R_n(\beta)\}=1$。

证明 只对$m=1$的情形加以证明，$m>1$可类似处理。只需证明

$$N_1\longrightarrow\infty,\quad n\longrightarrow\infty. \tag{9}$$

令

$$x_n=n^\omega,\quad\frac{\delta}{2}<\omega<1-\frac{1}{\kappa}-\frac{\delta}{2},\quad\frac{\kappa-1}{2\kappa}<\delta<\frac{\kappa-1}{\kappa},$$

$$C_n=\Big\{\max_{k=0,\cdots,2^j-1}\big|w_k(\xi)+w_k(e)\big|\le x_n\Big\},$$

分两步来证明。首先在 C_n 下证明 (9)，其次证明

$$\lim_{n\longrightarrow\infty} P(C_n) = 1. \tag{10}$$

(i) $H_1(1)$ 成立时，由引理 1 可知存在唯一的 $k_1 \in \{0,\cdots,2^j-1\}$，使得 $|w_{(1)}| = |w_{k_1}|$，且存在 $k_2 \neq k_1$，使得 $|w_{(2)}| = |w_{k_2}|$。利用 (6) 式有 $v_n = o(x_n)$，$x_n = o\big(n^{1-\frac{1}{\kappa}-\frac{\delta}{2}}\big)$，故

$$|w_{k_1}| \geq |w_{k_1}(f)| - x_n \geq n^{1-\frac{1}{\kappa}-\frac{\delta}{2}}(1+o(1)), \quad |w_{k_2}| \leq |w_{k_2}(f)| + x_n \leq v_n + x_n = O(x_n),$$

因此

$$N_1 = |w_{(1)}| - |w_{(2)}| \geq n^{1-\frac{1}{\kappa}-\frac{\delta}{2}}\big(1+o(1)\big) \longrightarrow \infty(n \longrightarrow \infty).$$

(ii) 证明 (10) 式成立。因为

$$P(C_n^c) = P\Big\{\max_{k=0,\cdots,2^j-1}\big|w_k(\xi)+w_k(e)\big| > x_n\Big\} \leq \sum_{k=0}^{2^j-1} P\big\{|w_k(\xi)+w_k(e)| > x_n\big\},$$

由引理 2、引理 3、假设 1 和稳定分布的尾概率性质可得

$$\lim_{n\longrightarrow\infty} P(C_n^c) \leq \lim_{n\longrightarrow\infty}\sum_{k=0}^{2^j-1} P\big\{|w_k(e)| > x_n\big\} \sim 2\lim_{n\longrightarrow\infty}\Big\{C_\kappa 2^j\big(2^{-j(\frac{1}{2}-\frac{1}{\kappa})}x_n\big)^{-\kappa}\Big\} = 0.$$

证毕

4 变点估计

N_i 不仅可以对变点存在性进行检验，还能够在变点存在时给出变点个数及位置的估计。设 q 为函数 f 的变点个数，变点位置依次为 $0 < \theta_1 < \cdots < \theta_q < 1$，$1 \leq q \leq m$，令 $\hat{q} = \sup\{i : 1 \leq i \leq m,\ N_i > c_i\}$，$k_1 < \cdots < k_{\hat{q}}$ 分别为 $|w_{(1)}|,\cdots,|w_{(\hat{q})}|$ 在排序之前所对应的指标，令 $(\hat{\theta}_1,\cdots,\hat{\theta}_{\hat{q}}) = (k_1/2^j,\cdots,k_{\hat{q}}/2^j)$，则有

推论 1 在定理 3 的条件下，当 $n \longrightarrow \infty$ 时

$$P(\hat{q} \longrightarrow q) = 1, \tag{11}$$

$$\sum_{i=1}^{\hat{q}}(\hat{\theta}_i - \theta_i)^2 = O_P\Big(\frac{1}{2^{2j}}\Big). \tag{12}$$

5 数值模拟

考虑如下模型

$$Y_i = f(X_i) + \lambda\varepsilon_i, \quad i = 1,2,\cdots,n, \tag{13}$$

其中 $X_i \sim U[0,1]$，$\lambda = 0.25$，ε_i 为特征指数 $\kappa = 1.83$ 的稳定过程。模拟分析为

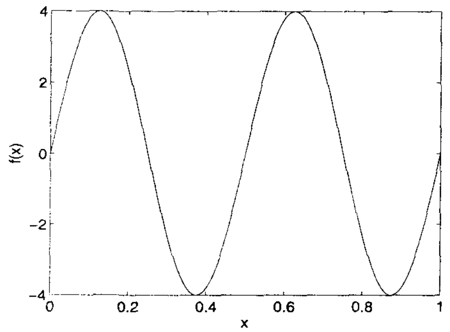
H_0: $f(x) = 4\sin(4\pi x)$;

H_1: $f(x) = 4\sin(4\pi x) - \mathrm{sgn}(x-0.3) - \mathrm{sgn}(0.58-x)$;

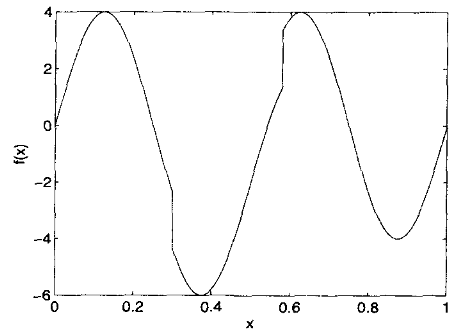
其中 H_0 中的 $f(x)$ 为 Wang[3] 所用的 heavisine 函数，见图 1(a)，H_1 中的 $f(x)$ 为含有两个变点的 heavisine 函数，见图 1(b)。图 1(c) 和 (d) 分别为 H_0 及 H_1 下数据 Y_i 的显示图。

取样本容量 $n = 2^{10} = 1024$，变点个数上界 $m = 3$，检验水平 $\beta = 0.05$。在分辨水平 $j = 5$ 下，用 Haar 小波对 Y_i 进行小波分解，得到的 H_0 及 H_1 下小波分解系数的绝对值 $|w_k|$ 分别见图 1(e)，(f)，其中水平虚线为本文给出的临界值 (critical region) 位置，水平实线是本文阈值 (threshold) 的最大值位置。

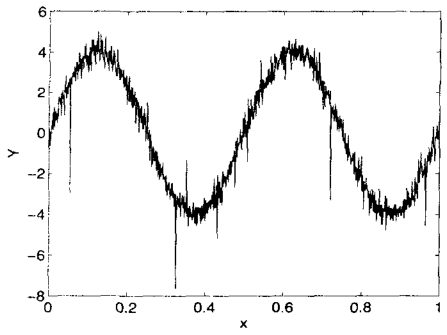
由图 1(e) 可以看出所有小波系数都小于本文给出的临界值，因此接受原假设 H_0；由 (f) 可以看出 $|w_k|$ 分别在 0.3 和 0.58 附近超过临界值，因此接受有两个变点的备择假设 H_1，这说明本文给出的检测与估计方法是有效的。



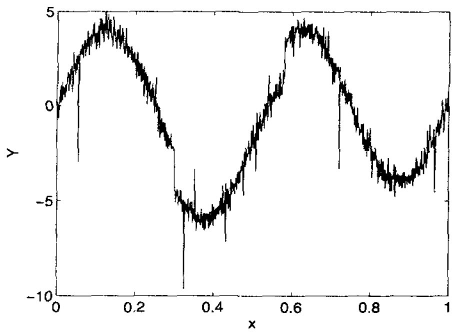
(a)



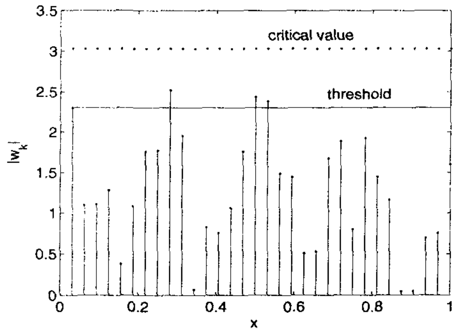
(b)



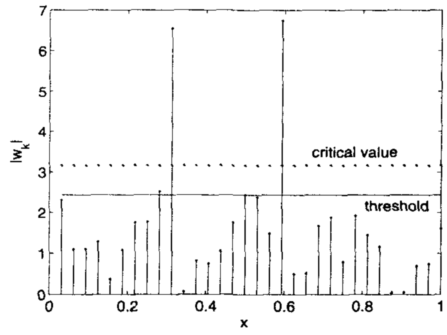
(c)



(d)



(e)



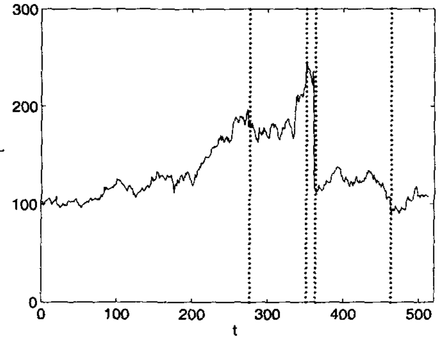
(f)

图 1: (a) 原假设 H_0 下 $f(x) = 4\sin(4\pi x)$； (b) 备择假设 H_1 下 $f(x) = 4\sin(4\pi x) - \mathrm{sgn}(x - 0.3) - \mathrm{sgn}(0.58 - x)$； (c) H_0 下模拟数据图； (d) H_1 下模拟数据图； (e) $j = 5$, H_0 下小波分解系数的绝对值； (f) $j = 5$, H_1 下小波分解系数的绝对值

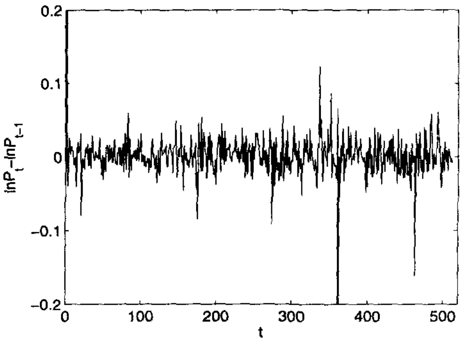
6　实例分析

本小节用美国IBM股票数据做实例分析，该数据为1997年12月19日至1999年12月31日512个IBM股票日收盘价，数据原始序列P_t见图2(a)。由图可知序列中存在跳跃点。因为股票数据常是厚尾序列，为此引入Mondebrot[7]中的方法估计特征指数κ。首先对原始数据进行对数差分见图2(b)，然后将$y_t = \ln P_t - \ln P_{t-1}$从大到小排列，最后用点$(\ln y_{(i)}, \ln(i/n))$拟合一条直线，直线的斜率便是$\kappa$的估计值，见图2(c)，估计得出$\kappa = 1.6235$，其中$i = 11, \cdots, 160$。利用本文的方法对该数据可能存在的变点进行检测与估计。

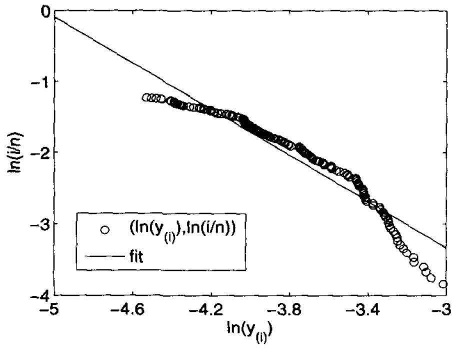
取变点个数上界$m = 6$，检验水平$\beta = 0.05$，在水平$j = 4$下，用Haar小波对数据进行小波分解得到的小波系数绝对值见图2(d)，其中的水平虚线是临界值位置。从图2(d)可以看出，小波系数在四处超过临界值，分别对应第274 (1999年1月22日)，336 (1999年4月22日)，361 (1999年5月27日)，和463 (1999年10月21日)，这说明在相应的时间附近有变点出现。图2(a)中的虚线给出的是变点实际发生的时刻。变点发生的部分原因可能是受到亚洲金融危机和俄罗斯金融危机及后续的影响。



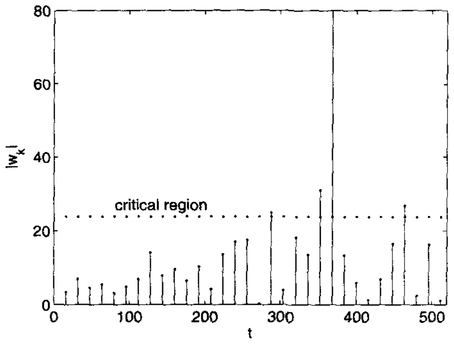
(a)



(b)



(c)



(d)

图 2:　(a) 512个IBM日收盘价原始数据图；(b) 原始数据对数差分图；(c) 特征指数估计图；(d) $j = 4$下小波分解系数的绝对值

参考文献:

[1] Carlstein E G, *et al.* Change Points Problems[M]. Hayward, Calif: Institute of Mathematical Statistics, 1994
[2] Csörgö M, Horváth L. Limit Theorems in Change-point Analysis[M]. New York: Wiley, 1997
[3] Wang Y. Jump and sharp cusp detection by wavelet[J]. Biometrika, 1995, 82: 385-397
[4] Odgen T, Parzen O. Change-point approach to data analytic thresholding[J]. Statistics and Computing, 1996, 6: 93-99
[5] Raimondo M. Minimax estimation of sharp change points[J]. The Analysis of Statistics, 1998, 26: 1379-1397
[6] Raimondo M. Tajvidi N. A peaks over threshold model for change-point detection by wavelet[J]. Statistics Sinica, 2004, 14: 395-412
[7] Mandebort B B. The variation of certain speculative prices[J]. Journal of Besiness, 1963, 36: 394-491
[8] Nolan J P. Stable Distribution-models for Heavy Tailed Data[M]. Boston: Birkhauser, 2004
[9] Mallat S. A theory for multi-resolution signal decomposition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11: 674-693
[10] Härdle W, *et al.* Approximation and Statistical Application[M]. Paris-Berlin: Springer-Verlag, 1998, Lecture Notes in Statistics, 129

Wavelet Estimation of Change-points in a Nonparametric Regression Function with Heavy-tailed Noise

LI Xiao-yan[1], TIAN Zheng[1,2], QI Pei-yan[1], ZHAO Wen-zhi[1]

(1- Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072; 2- Department of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710072)

Abstract: A wavelet method is proposed to detect and estimate change points in nonparametric regression models under random design, whose noise is a heavy tailed infinite-varianced process. If the null hypothesis holds, we obtain critical values at any scale and prove the consistency of the test by using the test statistics based on empirical wavelet coefficients. If the alternative hypothesis holds, we show that the estimation of the numbers and location of change points are consistent. Numerical simulation results as well as the IBM stock data analysis indicate that the proposed method is effective.
Keywords: random design; nonparametric regression model; infinite-varianced process; change points; wavelet estimation

Received: 21 Jan 2008. **Accepted:** 12 Nov 2008.
Foundation item: The National Natural Science Foundation of China (60972150; 10926197).